



## Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

4 | 2009

Linearization and Segmentation in Discourse

---

# A data-driven study of temporal adverbials as discourse segmentation markers

Lydia-Mai Ho-Dac and Marie-Paule Péry-Woodley

---



### Electronic version

URL: <http://journals.openedition.org/discours/5952>

DOI: 10.4000/discours.5952

ISSN: 1963-1723

### Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

### Electronic reference

Lydia-Mai Ho-Dac and Marie-Paule Péry-Woodley, « A data-driven study of temporal adverbials as discourse segmentation markers », *Discours* [Online], 4 | 2009, Online since 11 October 2017, connection on 16 December 2019. URL : <http://journals.openedition.org/discours/5952> ; DOI : 10.4000/discours.5952

---



*Discours* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

## **A data-driven study of temporal adverbials as discourse segmentation markers**

Lydia-Mai Ho-Dac, Marie-Paule Péry-Woodley

CLLE-Équipe de Recherche en Syntaxe et Sémantique, CNRS et Université de  
Toulouse UTM

### **Abstract:**

Many linguistic and psycholinguistic studies present sentence-initial temporal adverbials as “good” markers of discourse segmentation. This paper proposes a corpus-based evaluation of their potential to signal discontinuity such as a discourse break or shift. We look in particular at how temporal adverbials interact with other features, such as position in the text (in relation to document structure), and type of referring expression as grammatical subject in the host sentence. Our methodology calls upon a large diversified tagged corpus, and combines quantitative and qualitative approaches in order to systematically explore these configurations and their relation to text-type. Temporal adverbials are shown in this corpus to signal discontinuity only when in certain configurations, i.e., they cannot be considered as segmentation markers in their own right.

### **Keywords:**

discourse markers, discourse framing, discourse organisation, segmentation, temporal adverbials, referential accessibility, data-driven approach

### **Résumé**

De nombreuses études en linguistique et en psycholinguistique considèrent les adverbiaux temporels situés en position initiale de phrase comme de “bons” marqueurs de segmentation du discours. Cet article propose une analyse en corpus de leur capacité à signaler une discontinuité dans le discours (déplacement ou rupture). Nous portons ici une attention toute particulière aux configurations discursives dans lesquelles apparaissent les adverbiaux temporels à l'initiale en considérant deux autres traits : la position textuelle (relative à la structure du document) et le type d'expressions sujet apparaissant dans la même phrase que l'adverbial. Notre méthodologie se base sur un corpus diversifié et étiqueté, représentant un volume de données relativement important pour le domaine. Le va-et-vient entre analyses quantitatives et qualitatives nous permet d'explorer comparativement selon différents types de textes les configurations discursives auxquelles participent les adverbiaux temporels à l'initiale. Au vu de nos résultats, il semble que les adverbiaux temporels signalent une discontinuité uniquement dans certaines configurations, ce qui signifie qu'ils ne constituent pas en eux-mêmes des marqueurs de segmentation.

### **Mots-clés :**

segmentation, organisation du discours, adverbiaux temporels, encadrement du discours, marqueurs discursifs, accessibilité référentielle, analyse en corpus

## TEXTE INTEGRAL/ FULL TEXT

- 1 Segmentation has been a major research issue for discourse linguistics over time, as discourse has to conciliate complex informational structures with strong linearity constraints. Linguists are concerned with the definition of discourse units, the signalling of their boundaries and their relations; psycholinguists seek insights into the segmentation processes underlying text production and comprehension, and look at segmentation markers both as traces of text production and signals for the construction of a mental representation; computational linguists need to find operational ways of identifying meaningful segments, such as spans concerning a particular topic. Our approach is rooted in research in these three fields. Its original contribution lies in its systematic investigation of a large corpus of lengthy structured written texts. We focus on the role of sentence-initial temporal adverbials as segmentation markers, a role which has been the concern of a number of recent studies in the three fields mentioned above (e.g. for linguistics: Crompton, 2006; Le Draoulec, Péry-Woodley, 2003, 2005; Virtanen, 1992, 2004; for psycholinguistics: Bestgen, Vonk, 2000; Piérard, Bestgen, 2006; for computational linguistics: Bilhaut *et al.*, 2003, to quote but a few). The broader question we pursue in this paper is “what constitutes a discourse marker?”. Our empirical work leads us to question a narrowly lexical view of markers whereby specific expressions are seen as inherently endowed with a discourse role, and to propose instead a more complex view integrating different contextual parameters such as text-type and local textual strategies. Section 1 clarifies our take on textual segmentation, and surveys research into the role of adverbials in signalling segments. In section 2, we argue for a corpus approach and outline our largely data-driven method. Section 3 combines quantitative results and qualitative analysis to propose a strongly contextualised view of temporal adverbials and discourse segmentation.

### 1. Text segmentation

- 2 We take as our starting point the issue of sequentiality in text, as defined for example in Goutsos, who sees text as a “periodic alternation of transition and continuation spans” (Goutsos, 1996: 501). In his three-level representation of discourse structure, the cognitive level describes the writer's mental representation as structured by the basic strategies of continuity and discontinuity, while the linguistic level is concerned with the techniques available to realise these strategies, so as to produce, at the textual level, a segmentation into *continuation* and *transition* spans. The reader in turn uses knowledge of the instructional meaning of these signals to construct a mental representation. Our focus in this study is to uncover the instructional meaning of textual features which can play a role in signalling sequentiality.

#### 1.1. A matter of linearisation

- 3 Sequentiality in discourse can be seen as a consequence of the linearisation problem (Levelt, 1981): in text, a multidimensional discourse model is squeezed into a linear form. Linearity requires the writer to produce each textual unit in turn, and processing constraints demand short units of meaning. Yet the mental representation on which the discourse is based is not a succession of facts or ideas which can each be expressed in one sentence. This is where discourse organisation comes in, *via* two main relations: continuity and discontinuity.
- 4 Text segmentation has two related facets: (i) linguistic units cluster around a specific interpretation criterion; (ii) the text is divided into segments or spans. The interpretation criteria which bind text units together may have to do with discourse aboutness (e.g. referential continuity), or with organisation (e.g. rhetorical or document structure<sup>1</sup>). As long

---

<sup>1</sup> Cf. Power *et al.* (2003:211), who argue for an “abstract document structure” as a separate descriptive level, the purpose of which is “to mediate between the message of a text (i.e., its discourse

as a criterion remains valid, the segment is open and incoming linguistic units join into a continuation span. When it is no longer valid, the segment is closed and the resulting discontinuity is signalled via a transition span, which indicates a discontinuity in the discourse, e.g. a referential break, the end or start of a discourse frame, a rhetorical articulation, or the end or start of an element of document structure. Continuity being the default, a major task in the writing process is to signal discontinuity. In the absence of a cue to the contrary, the reader will interpret incoming sentences as continuous. Our concern is to identify these crucial cues which have the capacity to indicate a shift.

- 5 Some researchers argue that discontinuity is strongly signalled by specific cues – *segmentation markers* – whose function is described as indicating the boundaries of segments mostly defined in terms of thematic continuity. We propose a different definition, which rests on a different method of investigation. First, discourse organisation strikes us as typically signalled by the interaction of several cues rather than by specific markers. Next, we do not expect a one-to-one mapping between discourse cues and lexico-syntactic elements. Features such as attributes of document structure or text-type may play a part in a configuration. Finally, writers and readers have to manage several levels of organisation, which include thematic continuity but also temporal and spatial reference, rhetorical articulations and document structure. As discourse cues may be simultaneously contributing to several of these interdependent levels, a global view of discourse organisation is needed. Adverbials, in particular scene-setting adverbials, which are often associated with the expression of “clustering criteria” and therefore with discontinuity, are good candidates for a role in such configurations.

## 1.2. Structuring power, scope, and the segmenting capacity of scene-setting adverbials

- 6 Sentence-initial scene-setting adverbials are external to the clause, a syntactic independence which makes it possible for them to play an important role beyond their host clause or even sentence, in discourse organisation. We focus on sentence-initial adverbials because of the structuring power associated with elements in initial position. Starting point of the sentence, but also of a paragraph, section or text, the element in initial position influences the interpretation of what comes next in a different way for each level of organisation. In the case of adverbials, this influence is generally called *orientation* (Chafe, 1976; Firbas, 1986; Enkvist, 1976; Lambrecht, 1994; Dik, 1997). Orientation consists in setting the domain of applicability within which the following propositions hold. This link between orientation and adverbial use forms the basis of Charolles' *discourse framing* hypothesis (Charolles, 1997; Charolles *et al.*, 2005; Charolles, Vigier, 2005): sentence-initial adverbials acquire an instructional meaning relative to segmentation by projecting an interpretation criterion forward, and thus define the initial boundary of a segment. We distinguish two properties of sentence-initial adverbials which are called upon in this definition: *scope* and *structuring power*. *Scope* corresponds to the semantic continuity of the reference expressed by the adverbial. In example [1], the three sentence-initial adverbials (in bold) express time references which extend their respective scopes beyond their host sentences<sup>2</sup>:

[1]

### **FLORENCE-MILAN, 1500 - 1513** [heading]

**En 1500**, Léonard se rend à Mantoue, où il dessine le portrait d'Isabelle d'Este, [...], à Venise, [...], et à Florence, où - [...] - il va rester jusqu'en 1506. Son activité se partage entre des travaux de peinture : [...], et des travaux d'ingénieur militaire dans le val d'Arno et à Piombino. Léonard remet en chantier le Trattato commencé entre 1487 et 1492, et y travaille jusque vers 1513. **À partir de 1506**, il partage son temps entre

---

structure) and its physical presentation (i.e., its organisation into graphical constituents like sections, paragraphs, sentences, bulleted lists, figures, footnotes and so forth)."

<sup>2</sup> We provide an *English* translation for this first example in the text of the article; for all subsequent examples, English glosses can be found in the appendix.

Milan où [...], et Florence, où [...]. Il revient au projet de statue équestre, [...]. Il déploie une grande activité scientifique : anatomie, mathématique, et fournit des projets d'architecture, de décors pour Charles d'Amboise. **Mais, en 1513**, il quitte définitivement Milan reconquis par la coalition antifranaise.

**ROME-AMBOISE, 1513 – 1519** [heading]

À Rome, où il loge au Belvédère, Léonard se trouve [...]. [PEOPL]

[1'] (English version)

**FLORENCE-MILAN, 1500 - 1513** [heading]

In 1500, Leonardo goes to Mantova, where he draws Isabella d'Este's portrait, [...], to Venice, [...], and to Florence, where -[...] he stays till 1506. He shares his time between painting [...], and military engineering projects in the Arno valley and in Piombino. Leonardo resumes work on the Trattato started between 1487 and 1792, and continues until around 1513. From 1506, he divides his time between Milan where [...], and Florence where [...]. He returns to his equestrian statue project, [...]. He deploys an intense scientific activity: anatomy, mathematics, and produces architectural and decoration projects for Charles d'Amboise. But, in 1513, he leaves Milan for good as the city is reclaimed by the anti-French coalition.

**ROME-AMBOISE, 1513 – 1519** [heading]

In Rome, where he has his lodgings in the Belvedere, Leonardo finds himself [...]

- 7 We call *structuring power* (cf. Le Draoulec, Péry-Woodley, 2003, 2005) the capacity of initial adverbials to divide information into blocks which *structure* a (larger) span of text. Example [1] provides a clear example, with the three discourse frames conferring a temporal organisation to the entire section, while the topic remains Leonardo throughout. As well as creating temporal structure, these temporal frames also embody different geographical stages in Leonardo's life<sup>3</sup>
- 8 Viewed in the light of the discourse framing hypothesis, sentence-initial adverbials look like good segmentation markers as they introduce new discourse frames which can extend beyond the clause and structure a portion of text. A similar analysis can be found, though in different frameworks, in earlier studies such as Virtanen (1992, 2004) or Bestgen & Vonk (2000). We will at this point focus on two recent data-based studies which ask questions similar to ours, but use different methods and come to different conclusions.
- 9 Crompton (2006) offers an analysis for English which partly contradicts Charolles' hypothesis. Concerned with the scoping – rather than the structuring – properties of adverbials, Crompton concludes to the absence of any specific scoping potential linked to initial position. He argues that the scope of an adverbial is just a matter of default continuity. A setting will last as long as no new setting is introduced to cancel it. The author suggests the following processing principle: “Wherever (initially or non-initially) and however (adverbially or non-adverbially) a new circumstance is presented, apply the circumstance expounded to the next clause in the discourse unless cancelled” (Crompton, 2006:275-6). He ends on a view which is quite the opposite of Charolles': “the semantic function of positioning adverbials initially operates at a local, clause level rather than at a higher level within the discourse. Rather than providing an orientation for new discourse segments and thus making explicit higher-level text structure, positioning an adverbial initially serves to provide an orientation *for a single new clause*: in cases where comprehension of a particular circumstance depends on reference to previous discourse, the function of pre-posing of the adverbial would be *to facilitate the processing of the host-clause*” (*op. cit.*, our emphasis).
- 10 Piérard & Bestgen (2006), on the other hand, are concerned with the structuring properties of temporal adverbials. Following along the lines of Bestgen & Costerman (1997) and

<sup>3</sup> The heading provides an orientation towards such an interpretation.

Bestgen & Vonk (2000), they consider the segmentation marking function as primary because, in their view, an adverbial can only direct the interpretation of a text segment if it has already established the presence of a break. In order to evaluate the potential of temporal adverbials to indicate a break, they use two markers of topic shift which they consider reliable: paragraph break and a break in lexical cohesion (measured *via* the Latent Semantic Analysis method). They find that sentences beginning with temporal adverbials have a much higher chance of starting a paragraph<sup>4</sup> and correlate with lexical cohesion breaks. These observations lead them to conclude that temporal adverbials are good segmentation markers and therefore can reliably be associated with topic shifts.

- 11 Our own study questions the conclusions reached by both these papers, and provides new data to draw finer conclusions: temporal adverbials do act as segmentation markers, but only in specific configurations.

## 2. A corpus approach: rationale, methodology and results

- 12 Our research design aims to provide ways of measuring the contribution of different linguistic features in the signalling of sequentiality. It is part of a wider study which experiments with novel ways of exploring the signalling of discourse organisation in large corpora, focusing on the role of initial elements, i.e., all pre-verbal elements (Ho-Dac, 2007, 2008; Ho-Dac, Péry-Woodley, 2008). The methodology is governed by three basic principles: we study longer texts, where discourse organisation is necessary; we apply a quantitative data-driven approach; we start from the premise that discourse signalling results from an interaction between several forces and may also concern extra-linguistic features.
- 13 The first principle has to do with the impact of length on discourse organisation. Whereas it is possible for short texts to rely solely on referential continuity, longer texts, and in particular non-narrative texts with no default structuring in terms of succession of events, require other forms of organisation.
- 14 The second principle concerns our choice of a quantitative approach on a large corpus. Bestgen *et al.* (2006) point to the problems linked to the predominantly manual and qualitative nature of discourse research: weak representativeness, analyst-dependent annotations. We propose an alternative, with a largely data-driven methodology rooted in the hypothesis which forms the third principle: the signalling of discourse organisation is typically performed by multiple cues acting in synergy rather than by single discrete text elements. This synergy can be stated as follows: element A co-occurring with element B in a specific text-position in a particular text-type leads the reader to interpret either a continuity or a shift between preceding discourse and incoming information. We “play” with each kind of cue (text-type, text-position, element A and element B) so as to measure its influence on others.
- 15 Two of the features mentioned in our list of potential sequentiality cues are not lexico-syntactic: text-type and location (text-position). Text-type is a cue if we see it as working together with textual cues. Text-position is defined in terms of level of segmentation. We distinguish three text-positions related to document structure and associated with three levels of orientation: S1 = section-initial sentence; P1 = paragraph-initial sentence; P2 = sentences within a paragraph.
- 16 We use two measures to systematically explore these variation parameters in our search for configurations of cues: 1) deviations in the usage of different linguistic elements in initial position; 2) degree of association between two elements occurring in initial position: detached elements and grammatical subjects.

---

<sup>4</sup> A similar (but weaker) probability is also observed for sentences containing a temporal adverbial in median or final position.


## 2.1. Corpus description

- 17 To take into account textual variation, we constructed three corpora of French expository texts differentiated in terms of subject-matter and rhetorical organisation: ATLAS (~205,000 words), composed of 3 descriptive social geography texts; GEOPO (~250,000 words), a collection of 32 argumentative texts in the field of international relations; PEOPL (~220,000 words), 30 descriptive biographies. The texts in ATLAS are much longer than in GEOPO and PEOPL. They are mostly organised in terms of spatial and temporal locations acting as settings for large spans of text, with little topical continuity. Conversely, the texts in PEOPL are structured by strong topical continuity (the topic being the subject of the biography). Though all texts include sporadic temporal organisation, time as a structuring principle is not dominant and never extends to the whole text. The GEOPO corpus is more difficult to characterise, with occasional temporal organisation and rather weak or local topical continuities.

## 2.2. Annotation decisions and procedures

- 18 All pre-verbal elements in these three corpora were annotated automatically. Pre-verbal elements can be either syntactically detached or integrated. Detached elements occur before the grammatical subject, possibly – but not systematically – separated by a comma (e.g. adverbials, appositions, fronted adjuncts). Integrated elements correspond to grammatical subjects. Each detached element was annotated for part of speech, syntactic function, and semantic function in the case of scene-setting adverbials. Sentence-initial temporal adverbials represent 21% of all detached elements (1469 of the 7021 detached elements extracted). They are equally distributed in the three corpora: 31% come from ATLAS, 35.5% from GEOPO and 33.5% from PEOPL.
- 19 Grammatical subjects are automatically placed on a scale of cognitive accessibility adapted from Ariel's Theory of Accessibility (1990), on the basis of part of speech tag, presence of a lexical reiteration, and length (Table 1). All these features are automatically detected by a set of regular expressions based on information obtained from automatic POS tagging (Treetagger<sup>5</sup>) and syntactic dependency analysis (Syntex – Bourigault, 2007).

**Table 1. Scale of seven degrees of accessibility adapted from Ariel's Theory of Accessibility (1990)**

indefinite description or special construction	0		low acc.
proper name without lexical reiteration	1		
long definite description without lexical reiteration	2		
definite description with lexical reiteration or short	3		
reiteration of a proper name ("redenomination")	4		
long demonstrative description without lexical reiteration	5		
demonstrative description with lexical reiteration or short	6		
pronoun or possessive NP	7		high acc.

- 20 Accessibility models such as Ariel's (1990) (cf. also Gundel *et al.* (1993), Centering Theory (Walker *et al.* 1998)) are not designed to classify referential expressions in terms of continuity or discontinuity: they aim to account for the cognitive processes involved in the activation of discourse referents. There is nevertheless a potential correlation between activation of discourse referents and topical continuity. As stated by these models, one way to keep a referent activated is to use morpho-syntactic elements that are associated with a

<sup>5</sup> <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

high degree of accessibility (e.g. pronouns), whereas the introduction of a new referent must be performed via elements associated with low accessibility (e.g. indefinite NPs). In this study we consider grammatical subjects with a high degree of accessibility as potential continuity cues and grammatical subjects with a low degree of accessibility as potential discontinuity cues.

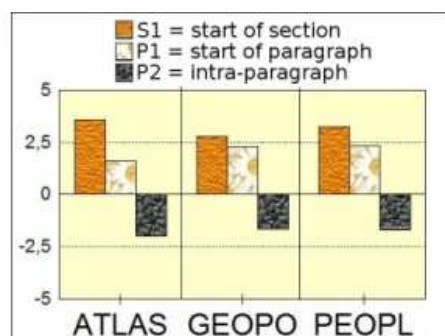
- 21 From these data, we extracted configurations for each text type and text-position by measuring the following variations for the different features: (i) distribution in each corpus compared with overall distribution; (ii) distribution in each text-position compared with overall distribution. The significance of variations is given in terms of z-score. We regard as significant a deviation above or below 2.5.
- 22 Our systematic corpus annotation now makes it possible to take a quantitative look at the configurations of cues where temporal adverbials appear. These configurations are described along the dimensions defined previously: text-type, text-position, and preferred associations (co-occurrence) between detached elements and referring expressions occupying grammatical subject function. On the basis of these parameters, and through an approach which combines large-scale quantitative aspects and qualitative zooms, we are able to gain new insights into the question of the structuring power of temporal adverbials. Some of the data below are set out as frequencies, but we mostly present our results in terms of z-scores to bring out deviations from the theoretical norm (i.e., whether the objects under study are evenly distributed throughout the data considered).

### 3. Context-sensitive temporal adverbials<sup>6</sup>

#### 3.1. Temporal adverbials in three text-positions in three corpora

- 23 Figure 1 shows the deviations for sentence-initial temporal adverbials in the three text-positions defined in section 2: S1 (start of section), P1 (start of paragraph) and P2 (intra-paragraph). It reveals in all corpora a marked preference of these adverbials for the S1 position (z-score  $>+2.5$ ), and to a lesser degree for P1.

**Figure 1. Sentence-initial time adverbials: deviations acc. to text-position in three corpora**



- 24 The intra-paragraph position is generally dispreferred (though not significantly so). It must be noted that this distribution is not shared by other types of adverbials, which should make us wary of generalisations. For example, space adverbials only show significant variation in P1 and in ATLAS, as Figure 2 illustrates.

<sup>6</sup> *Adverbial* in this section is to be taken as meaning *sentence-initial adverbial*, even when this is not specified (cf. section 2).



**Figure 2. Sentence-initial space adverbials: deviations acc. to text-position in three corpora**



- 25 Indeed, despite the regularity of the distributions in Figure 1, delving deeper into our results will lead us to stress the variability of adverbial usage. Nevertheless, this global view points to a clear correlation between document structure (sections and paragraphs) and occurrence of sentence-initial temporal adverbials. We suggest that positions S1 and P1 are important parameters in the orientation function these expressions can fulfil.
- 26 It may be useful at this juncture to return to Crompton's study, even though it encompasses a much broader object, with initial adverbials in general as its concern. To rephrase this concern in terms of our narrower interest: sentence-initial temporal adverbials are associated in the literature with two discourse functions or meanings, namely i) extending their scope over larger discourse spans (Crompton's Meaning 1); ii) signalling boundaries between spans of discourse, or units of discourse structure (Crompton's Meaning 2). The author "assume[s] that Meaning 2 is derivative from Meaning 1, with the common factor of being under the scope of the same adverbial helping to constitute discourse spans." (Crompton, 2006:249). His method for testing this assumption is limited to the analysis of the scoping behaviour of adverbials, i.e., Meaning 1, because, in his words "to investigate Meaning 2 independently would require a fairly robust model of discourse structure, and I do not feel that such a model exists yet" (Crompton, 2006:250). We do not dispute his view regarding the availability of a mature model of discourse structure, but we will argue that our empirical method is capable of providing some elements of answer. Using insights from our wider study, we will call upon the related notions of discontinuity, structuring power and scope introduced in section 1.2. Negative results obtained by Crompton could be linked to the fact that Crompton's study focuses strictly on scope and not on structuring power. As Le Draoulec & Péry-Woodley (2003; 2005) show, the semantic scope of temporal adverbials in narrative texts is generally limited to one clause. Adverbials devoid of semantic scope – whatever the reason, e.g. conflict with advancing time reference in narration – can however still be endowed with structuring power. Moreover, temporal adverbials may indicate something other than temporal organisation, such as a rhetorical articulation, a topic shift or, as in Example [1], a spatial organization. There would therefore appear to be a degree of independence between the structuring power of an expression and its semantics.
- 27 Another consideration which may account for Crompton's negative results in his search for scoping properties is his choice of texts. His study is based on a corpus exclusively made up of very short texts (editorials and student essays). In such short texts (*circa* 500 words) writers simply do not need structuring devices beyond referential continuity. In long texts on the contrary, writers tend to resort to a variety of different devices.
- 28 Whereas the initial findings in Figure 1 do not square with Crompton's analysis, they largely concur with findings from earlier psycholinguistic and corpus studies (Bestgen, Vonk, 2000; Piérard, Bestgen, 2006). The question pursued by Piérard & Bestgen is however somewhat different. Their interest, as we saw in section 1.2, is in evaluating the potential of temporal adverbials to indicate topic shifts. They use two markers of topic shift they consider reliable – paragraph break and a break in lexical cohesion – to evaluate the

segmentation potential of certain lexical expressions, here of temporal adverbials. The observation that sentences containing temporal adverbials have a much higher chance of starting a paragraph and that this chance increases when the temporal adverbial occurs in sentence-initial position leads them to the conclusion that such adverbials are good segmentation markers and therefore can reliably be associated with topic shifts. We will comment later on the assumed link between discontinuity and topic shift. For now, our focus is on the following question: are temporal adverbials intrinsically good discontinuity markers, or do they simply “ride” on the segmentation potential of the paragraph break? This strikes us as an important question for linguistics and for computational linguistics. In linguistic terms, the traditional essentially lexical view of *discourse markers* is at stake. For computational linguistics, the correct identification of segment boundaries is obviously a major objective.

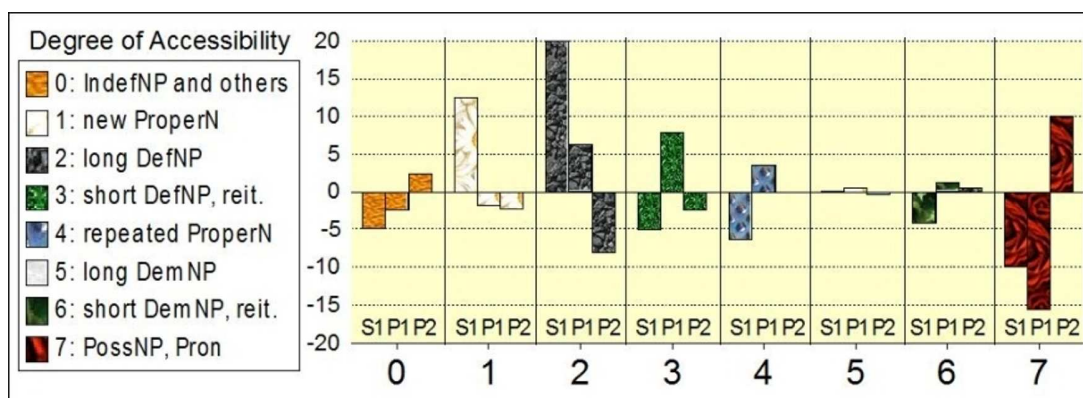
### 3.2. Putting the structuring power of temporal adverbials to the test

- 29 In order to test the structuring power of the adverbial itself when it coincides with a new paragraph, we examine its environment in other text-positions, i.e., P2. If adverbials in intra-paragraph positions turn out to regularly co-occur with discontinuity markers, such as referring expressions of low or medium accessibility (suggesting thematic discontinuities), this could be taken as indicating that these adverbials have in themselves the potential to signal a discontinuity. Conversely, the absence of significant variation could mean that their role as discontinuity markers is dependent on other parameters, yet to be identified, amongst which a new paragraph or section. This test can initially be formulated in the form of two questions:

- what referring expressions do P2 temporal adverbials co-occur with in their host sentence? If they do signal discontinuity, we should observe preferred associations with new referents (corresponding with a degree of accessibility of 0, 1 and 2) or lexical reiterations (corresponding with a degree of accessibility of 3 and 4).
- what referring expressions do P2 temporal adverbials co-occur with in the sentence following their host sentence? if temporal adverbials open a new segment, we can expect to observe continuity in what follows within this new segment, i.e., preferred associations with high degrees of accessibility.

- 30 To allow informed comparison, Figure 3 shows general referring expression usage.

**Figure 3. Degree of accessibility: deviations acc. to text-position in the entire corpus**



- 31 Three groupings come out in Figure 3: grammatical subjects with a degree of accessibility (abbreviated as *DegAccess*) of 1 and 2 occur significantly more in S1, *DegAccess* 3 and 4 subjects in P1, and *DegAccess* 7 subjects in P2. These data are in agreement with a number of linguistic descriptions which can be summarized as follows:

- S1 position (start of a section) usually coincides with a break i.e. re-initialization of referents and settings in a new rhetorical move;

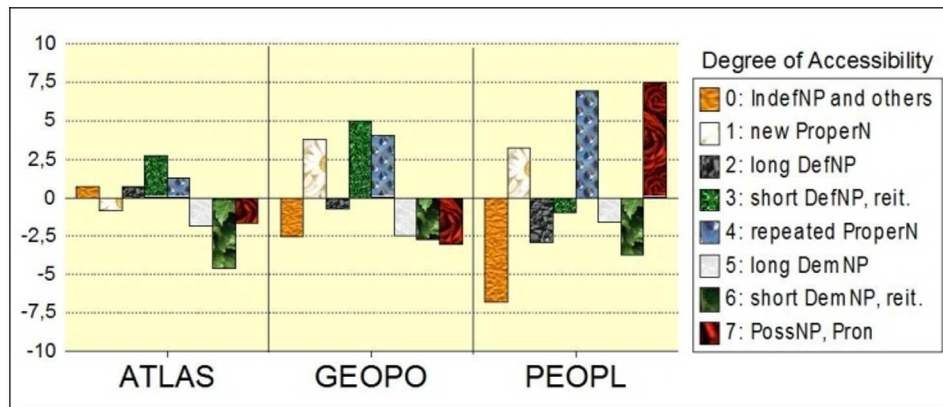
- P1 position (start of a paragraph within a section) usually coincides with a shift i.e., a change of setting or a rhetorical move which requires reiteration of current referents - when the text is organised around referential continuity -; or conversely a change of referents while keeping the same setting within a move;
- in P2 position (intra-paragraphic), default referential continuity applies, allowing the use of expressions with little referential information such as pronouns.

32 We will comment below on our use of a cognitive accessibility scale, and on the somewhat rough and ready association between types of referential expression and degree of accessibility. Suspending judgment on these methodological choices for the time being, we are able to distinguish three kinds of grammatical subjects: those linked to breaks, those linked to shifts and those linked to continuity. The structuring power of temporal adverbials can now be tested in terms of these correlations.

### 3.2.1. Co-occurring referring expressions in the host sentence

33 Figure 4 shows the deviations in degree of accessibility of grammatical subjects in sentences starting with a temporal adverbial, the reference being all sentences for each corpus. The intra-corpus variability is considerable, with PEOPL showing the greatest deviations from the “norm”: the strong association between temporal adverbials and DegAccess 7 (possessive NPs and pronouns) is the exact opposite of what the hypothesised segmentation potential of temporal adverbials would predict. The other strong association in PEOPL is with Proper names (new (DegAccess 1) and repeated (DegAccess 4)), a tendency shared, to a lesser extent, by GEOPO, where the preferred association is with DegAccess 3 (definite descriptions). GEOPO and PEOPL significantly disprefer indefinite descriptions and special constructions together with a temporal adverbial, while ATLAS remains largely indifferent.

**Figure 4. Deviations in DegAccess in sentences with temporal adverbials w.r.t. all sentences in the three corpora**



34 The absence of significant variation in ATLAS shows that temporal adverbials do not disrupt the general model for referential expressions. This could be related to the strong overall spatio-temporal organisation of this corpus where topical continuity is rare. The dominant referring expressions are long definite descriptions with lexical reiteration (e.g. *the proportion of workers' children schooled in the public education system*), i.e., DegAccess 3. In this corpus, the majority of temporal adverbials occur in section-initial sentences (see the strong deviation in figure 1). The adverbials' structuring and scoping properties can be seen to be strongly related to document structure. Example [2]<sup>7</sup> illustrates this relation between temporal organisation and document structure in ATLAS. Note that such associations could not be found in short texts such as were used by Crompton to test adverbial scope.

<sup>7</sup> An English gloss is provided in the appendix, as for subsequent examples.

[2]

## 5. LE COLLÈGE [heading]

Dans les années 1970, ce sont les problèmes des collèges qui ont tenu le devant de la scène de l'actualité scolaire: effectifs trop nombreux, inadaptation des enseignements à la diversité des élèves, exclusion des laissés-pour-compte d'un enseignement élitiste. Ces difficultés n'ont pas disparu, bien qu'aujourd'hui elles soient plus sensibles au lycée.

### 5.1. L'INÉGALE CROISSANCE DES EFFECTIFS DU SECOND DEGRÉ [subheading]

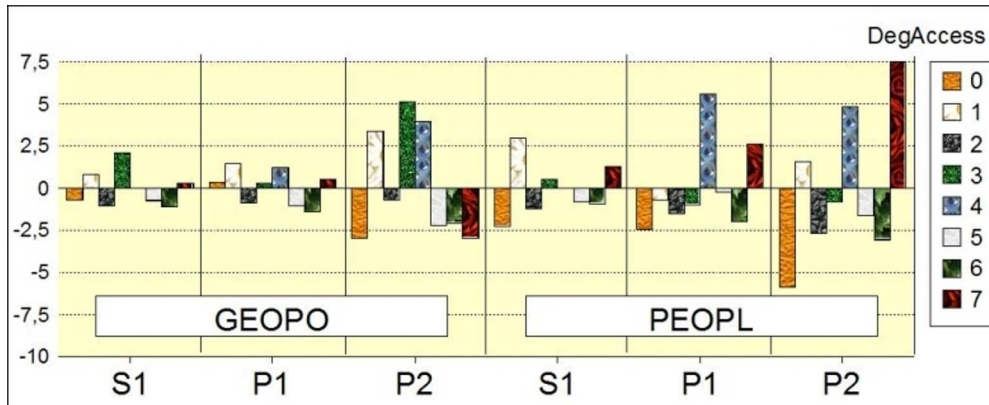
À la rentrée 1965-66, on comptait 3 250 000 élèves dans les collèges et lycées. Vingt ans plus tard, 5 100 000; et plus de 5 500 000 à la dernière rentrée. La secondarisation des études et la poursuite jusqu'en Terminale de lycée par plus d'un jeune sur deux constituent le changement le plus fondamental survenu depuis les années 1960 au sein du système éducatif français. Mais ce changement a été inégal selon les régions, si bien que subsistent des disparités, comme autant de déphasages dans le processus général de secondarisation des études.

De 1965 à 1985, le nombre de collégiens et de lycéens a augmenté de 70%, mais selon des rythmes et avec des intensités différents selon les académies et les départements. Faible dans le Sud-Ouest et le Massif central, modérée en Bretagne et à Paris, l'augmentation a été considérable dans le Centre-Ouest, en Alsace, dans la région Rhône-Alpes et dans les départements de la grande banlieue parisienne où les effectifs ont souvent plus que doublé. Les variations de la population et les baisses plus ou moins fortes du nombre des naissances selon les régions ne suffisent pas à expliquer ces différences d'accroissement des effectifs du secondaire. Intervient aussi l'allongement des scolarités, qui a été plus marqué dans les départements où, au milieu des années 1960, la poursuite des études après l'école primaire était loin d'être la règle.

À la rentrée 1990-91, les 6 700 collèges de France métropolitaine ont accueilli 3 100 000 élèves, chiffre en diminution de plus de 300 000 sur celui de l'année scolaire 1985-86. Les 1 800 collèges privés scolarisent un peu plus de 20% des élèves. Leur part, qui s'est accrue de deux à trois points au cours de la dernière décennie, est surtout forte, comme pour le primaire, dans la France de l'Ouest, les départements du sud et de l'est du Massif central, à Paris et dans le Nord. [ATLAS]

- 35 Having set the general picture with Figure 4, let us turn to Figure 5, where the data are now broken down according to text-position: S1, P1, P2. We now have the means to evaluate the impact of temporal adverbials independently of paragraph breaks, i.e., in P2 position. As the data for ATLAS are not particularly telling for the matter in hand, and in order to simplify a somewhat complex argument, we have opted to leave ATLAS out and focus on GEOPO and PEOP.

**Figure 5. Deviations in associations between a temporal adverbial and DegAccess n w.r.t. the general model for each text-position in two corpora**



- 36 Figure 5 shows the extent to which the significant associations observed in Figure 4 are irregularly distributed among text-positions. In GEOPO, they are all to be found in intra-paragraphic sentences (P2), in other words GEOPO follows the general model for S1 and P1, (like ATLAS), i.e., break in S1 and shift in P1 (cf. Fig. 2). All the significant P2 deviations in GEOPO suggest a correlation between the presence of a temporal adverbial and discontinuity, as breaks tend to be linked to DegAccess 1 and shifts to DegAccess 3/ 4.
- 37 In PEOPL, it becomes clear that the association with DegAccess 7, possessives NPs and pronouns, is largely accounted for by P2 (z-score=7.5). Repeated proper names (DegAccess 4) are also strongly represented in P2 and P1. These associations in PEOPL can now be seen to agree with the general model: low degree of accessibility in S1 (section-initial), indicating a break<sup>8</sup>; medium in P1 (paragraph-initial), indicating a shift; finally in P2 (intra-paragraph) medium, and most remarkably highest degree of accessibility, indicating topical continuity.
- 38 Two examples from PEOPL will throw qualitative light on these matters (temporal adverbials are in bold and grammatical subject in bold italicized). Texts in these biographies naturally centre on one topical referent, yet the absence of topic shift does not invalidate the structuring power of temporal adverbials in positions S1 and P1, particularly in a configuration we call after Virtanen temporal Text Strategic Continuity (TSC, Virtanen, 1992, 2004). Example [3] (repeating example 1) shows such a string of adverbials, structuring the paragraph into three time periods.

[3]

**FLORENCE-MILAN, 1500 - 1513** [heading]

**En 1500, Léonard** se rend à Mantoue, où il dessine le portrait d'Isabelle d'Este, [...], à Venise, [...], et à Florence, où - [...] - il va rester jusqu'en 1506. Son activité se partage entre des travaux de peinture : [...], et des travaux d'ingénieur militaire dans le val d'Arno et à Piombino. Léonard remet en chantier le Trattato commencé entre 1487 et 1492, et y travaille jusque vers 1513. **À partir de 1506, il** partage son temps entre Milan où [...], et Florence, où [...]. Il revient au projet de statue équestre, [...]. Il déploie une grande activité scientifique : anatomie, mathématique, et fournit des projets d'architecture, de décors pour Charles d'Amboise. **Mais, en 1513, il** quitte définitivement Milan reconquis par la coalition antifranaise.

**ROME-AMBOISE, 1513 – 1519** [heading]

<sup>8</sup> Our annotation does not account for reiteration in S1. This bias does not however affect the present analysis.

À Rome, où il loge au Belvédère, **Léonard** se trouve [...]. [PEOPL\_Léonard de Vinci]

- 39 This strategy is initiated in the section heading and reinforced by the section-initial adverbial *En 1500*. Numerous examples like [3] account for the highly significant preference for DegAccess 7 forms (*il, he*) after a P2 adverbial. We also observe a propensity for these P2 adverbials to be followed by repeated Proper names (*Léonard*), as in example [4], where pronouns and reiterations of the protagonist's name alternate. Schnedecker (2005) notes this particular usage of repeated proper names in portraits of personalities in the written press: in such texts with high topical continuity, the repeated proper name loses its redomination impact (cf. Centering Theory *inter alia*), and works largely as an alternative to the anaphoric pronoun. This qualitative analysis pulls repeated proper names in these texts towards the high accessibility end of the scale.

[4]

### 1.3) LE PROFESSEUR ET LE FOLKLORISTE [heading]

**En janvier 1907, Béla Bartók** est nommé professeur [...]. Cette nomination le met à l'abri des soucis matériels. **Il** fait consciencieusement son enseignement **Il** poursuit [...]. Son attachement pour Strauss dépassé, **il** découvre la musique de Debussy [...]. Le message [...]. De 1907 date un Concerto [...]. 1908 marque le début [...]. Son échec sentimental courageusement surmonté par le travail, **Bartók** épouse, en novembre 1909, son élève Márta Ziegler. Le couple fait son voyage de noces à Paris,... **En mars 1910, il** est de nouveau à Paris [...]. Une semaine plus tard, le jeune Quatuor [...]. **À la fin de mai, Bartók** joue sa Rhapsodie [...] à Zurich [...]. **À partir de 1912**, découragé par le non-intérêt de la société hongroise pour la musique nouvelle, **Bartók** se retire de la vie musicale, [...]. ; **il** poursuit toutefois [...] [PEOPL\_Bartok]

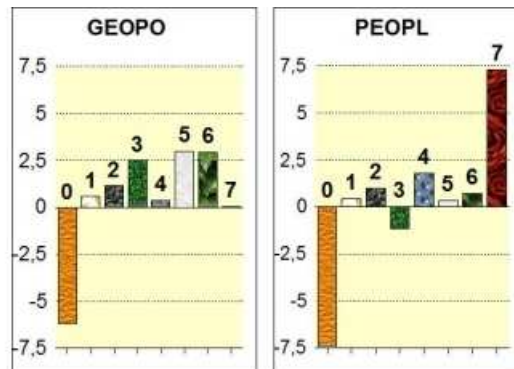
- 40 Examples [3] and [4] show the dominant organisation in PEOPL to be topical continuity, the temporal “stepping stones” having no real extended scope. This should not however be seen as meaning that they are unable to structure text (cf. the structuring power of temporal adverbials in [3]); but in terms of our first question, we find no evidence for an association between temporal adverbials and low DegAccess referring expressions in PEOPL. The opposite holds true for GEOPO, where our results agree with those of Piérard & Bestgen (2006).
- 41 We next address our second question, concerned with continuity in the putative segment opened by a P2 temporal adverbial, as evidenced by referring expressions used in what follows the adverbial's host sentence.

### 3.2.2. Co-occurring referring expressions in the next sentence (Sentence+1)

- 42 Figures 6 relates to the types of referring expression found in Sentence+1, i.e., the sentence immediately following a temporal adverbial's host sentence not coinciding with a paragraph break.



Figure 6. Deviations in DegAccess in Sentence+1 w.r.t. the model for GEOPO (P2) and PEOP (P2)



- 43 For GEOPO, Figure 6 shows a significant deviation for DegAccess 3, 5 and 6 from the general model (Fig 6: z-score +2,5, +3 and +2.9). The deviation for DegAccess 3 (short definite descriptions) indicates a shift in temporal reference rather than continuity (cf Fig. 2). Deviations for DegAccess 5 and 6 indicate another phenomenon: though Demonstrative NPs (DegAccess 5 and 6) are co-referential, their use in preference to a pronoun is often associated with “reclassification”, whereby a known referent is re-introduced stripped of its initial circumstances (De Mulder, 1997). This is typically what happens in GEOPO, as illustrated in [5]: a new referent, *des écoles* (schools), is introduced after an adverbial (*Lors des bombardements de 1998*; *During the 1998 bombings*) which could be opening a discourse frame. But the referent's reclassification in the next sentence (*Ces écoles*; *These schools*) negates the possibility of an extension of the adverbial's scope. The adverbial's function seems to be to introduce the new referent, and its scope does not extend beyond the host sentence:

[5]

#### UN SCÉNARIO POSSIBLE POUR UNE GUERRE ANNONCÉE [heading]

La plasticité du régime [de Saddam Hussein] est un facteur ignoré dans toutes les anticipations de la guerre. [...] Les "frappes" et autres ingérences étrangères l'ont préparé à cette confrontation ultime. Elles lui ont appris à escamoter ses cibles [...] Le régime escamote parfois jusqu'aux cibles les plus ordinaires. **Lors des bombardements de 1998, des écoles**, ainsi que des installations industrielles et des hangars alimentaires, ont accueilli des dépôts de munitions. **Ces écoles** abritent actuellement les membres du Parti chargés de maintenir l'ordre dans chaque quartier. Ceux-ci ont quitté leurs locaux officiels, [...]. [GEOPO]

- 44 Whether an initial adverbial has an extended scope seems to depend very much on its text-position. At the start of a section or a paragraph, it tends to extend its scope over the unit it initiates, unless there is an indication to the contrary (such as a new temporal location). In example [6] such an adverbial (*En 1993*) is seen to extend its scope until the appearance of a different temporal location in the first sentence of the next paragraph:

[6]

#### LA BOSNIE [heading]

**En 1993**, les débats sur la possibilité d'un engagement militaire en Bosnie se sont accompagnés de nombreuses critiques sur le coût des opérations, la nécessité d'engager les forces armées et la participation des alliés européens. Les parlementaires se sont interrogés sur les pouvoirs de guerre accordés au président Bill Clinton [...]. Certains membres du Congrès se sont montrés sceptiques quant aux prétentions de la Maison-Blanche à jouer les "gendarmes du monde". Ils estimaient que le pays, en intervenant de façon excessive à l'extérieur, risquait de gaspiller ses ressources et de dévoiler trop

aisément ses forces à ses adversaires. Le sénateur John McCain (républicain, Arizona) pensait ainsi que, "si nous usons de nos forces et de notre prestige de façon inconsidérée, nous gaspillerons des ressources que nous n'avons pas".

Des remarques du même type ont été exprimées **en 1995**, notamment de la part d'élus démocrates, qui jugeaient que l'Administration avait prêté une trop grande attention aux aspirations de l'aile conservatrice du Congrès en décidant d'envoyer des troupes dans les Balkans. Le sénateur Byron Dorgan (démocrate, Dakota-du-Nord) estimait ainsi que "[...]". Pour sa part, et conscient de l'importance de la participation financière que supposait une intervention armée, le représentant Jerry F. Costello (démocrate, Illinois) considérait que la Bosnie concernait les Européens au premier chef et que ceux-ci devaient en assumer la principale responsabilité, écartant ainsi le principe d'un envoi de troupes américaines sur le terrain. [GEOPO]

- 45 Returning to PEOPL in Figure 5, we note a strong deviation for DegAccess 7. There are more pronouns or possessive NPs in the sentence following a temporal adverbial's host sentence than in the general model (cf. examples [3] and [4]). In other words, topical continuity is maintained after a temporal adverbial. But, as was seen in the preceding section, the presence of a temporal adverbial cannot be seen as responsible for this topical continuity since DegAccess 7 is also dominant in the host-sentence. A last example shows the clear dominance of topical continuity over temporal organisation in PEOPL. This dominance is indicated by systematic co-occurrence of temporal adverbial and pronoun in P1, and is reinforced by the narrative character of this section. As a result, the scope of the temporal adverbials is difficult to determine: in the third paragraph in particular, time references move very gradually from *En août 1901* to *à la fin d'août 1902* (from *In August 1901* to *in late August 1902*), followed by the co-referring temporal adverbial *Pendant cette année* (*In the course of that year*).

[7]

#### 1.1) LA JEUNESSE [heading]

Son enfance et son adolescence ont pour décor l'Italie, puis la Côte d'Azur et Monaco, où il réside de 1886 à 1898 avec sa mère et son frère, né deux ans après lui.

**En 1899, il** est à Paris avec sa famille : existence difficile, qui ne l'empêche pas d'écrire et de fréquenter les bibliothèques. L'été de 1899 le conduit pendant quelques semaines avec son frère à Stavelot, dans les Ardennes belges : le jeune poète découvre là un paysage et un folklore dont la trace se perpétuera dans son œuvre ; il y éprouve aussi une déception sentimentale dans laquelle on peut voir sa première expérience amère de "mal-aimé".

**En août 1901, il** part pour la Rhénanie comme précepteur. Il passera en Allemagne une année décisive dans l'élaboration de son univers mental comme dans sa vie amoureuse. Il est séduit par le pittoresque de la vallée du Rhin et de la région avoisinante, qu'il parcourt en tous sens. Il fait aussi, à la fin de l'hiver et au printemps de 1902, un long voyage à travers l'Allemagne et l'Europe centrale, s'arrêtant notamment à Berlin, Prague, Vienne et Munich. D'autre part, épris de la gouvernante anglaise de son élève, Annie Playden, il se voit éconduit par la jeune fille après avoir espéré lui faire partager son amour : c'est profondément désemparé qu'il rentre en France **à la fin d'août 1902**.

**Pendant cette année, il** n'a pas seulement vécu intensément. Il a écrit de nombreux poèmes, des contes, des chroniques, s'est nourri d'images et de sujets qui ne le quitteront pas. C'est pendant cette année aussi que paraît dans *La Revue blanche* de mars 1902 son premier conte, " L'Hérésiarque ", signé du pseudonyme qu'il s'est forgé depuis longtemps déjà avec deux de ses prénoms : Guillaume Apollinaire. [PEOPL]



#### 4. Conclusion: unstable structuring power

- 46 The picture we get from our combination of quantitative and qualitative approaches is quite complex. To summarise it, let us return to our two questions aiming to test the intrinsic structuring power of temporal adverbials. Do we find strong associations between temporal adverbials in intra-paragraph position and referring expressions with a *low* degree of accessibility? There is immediate evidence to the contrary in PEOPL with the clear preference for the highest degree of accessibility (DegAccess 7), but the high z-scores for medium accessibility expressions (DegAccess 3 and 4) require a careful qualitative exploration. The frequent repeated proper names in the intra-paragraphic sentences in PEOPL cannot be interpreted as strong redenominations (indicating a shift), but as an alternative for DegAccess 7 expressions, thereby supporting our initial results.
- 47 Our second question concerns the referring expressions in the sentence following the adverbial's host sentence: do we find a preference for referring expressions with a *high* degree of accessibility, indicating continuity within the putative segment opened by the adverbial? Here again, the answer is no, but again it requires adding fine linguistic analysis to the quantitative approach. The fairly marked preference for DegAccess 5 and 6 referring expressions in GEOPO could be seen as a positive answer to the continuity question, yet a close look reveals that these demonstrative descriptions, though they are of course co-referential, regularly indicate in this corpus a reclassification and therefore a shift in time reference.
- 48 We are aware of the piecemeal nature of our presentation of these analyses, and we certainly do not claim to have answered the question of the structuring power of adverbials. Our main objective here is to provide evidence of the importance of non-lexical parameters in the discourse role of expressions which are generally considered in and for themselves. Our data seem sufficient to justify two assertions regarding sentence-initial temporal adverbial. The first rephrases Charolles' parallel between paragraphs and discourse frames (Charolles, 1997): a paragraph is an underspecified segment, an initial temporal adverbial turns it into a specified segment, a discourse frame. Taking it one step further: an initial adverbial which does not coincide with a paragraph break does not have structuring power, and does not in itself act as a discontinuity marker.
- 49 Our second assertion concerns variability linked to text-type: quantitatively, the differences are immediately apparent; zooming in on a few examples shows how the relation between accessibility and continuity varies according to text type, an effect, among others, which inevitably induces variability in the discourse function of adverbials.
- 50 These two assertions put Crompton's results into perspective. Like Crompton's, our data suggest that sentence-initial temporal adverbials do not have intrinsic scoping potential. But in contrast with his study, our methodology allows us to show that the semantic role often associated with initial position can concern higher levels of discourse organisation if there is congruence with document structure. Such descriptive results require that multiple cues should be considered in interaction, and that research should not restrict itself to short texts, where paragraph breaks are absent or rare because unnecessary.
- 51 The unstable relation between accessibility, thematic continuity and temporal organisation questions the very definition of discourse discontinuity. Piérard & Bestgen (2006) seem to establish a clear link between a discourse break and thematic discontinuity. Our results suggest a more complex picture, with different types of discontinuity corresponding to different modes of organisation. It will also be necessary, in future work, to examine critically the use we make of Ariel's accessibility scale in order to characterise thematic continuity: it relies on an assumed but untested link between referential continuity and degree of accessibility. As was seen in the case of demonstratives or repeated proper names, the data must be interpreted with great care, in particular with middle degrees of accessibility (from 3 to 6), because the different co-referential expressions involved are

clearly concerned with more than just accessibility. Recent psycholinguistic studies (e.g., Fossard 2006) will be helpful in investigating the cognitive processes involved in the use of the different co-referential expressions.

- 52 This study also highlights the sensitivity of discourse organisation to text-type, both in terms of social genre and in terms of the document's shape (text length, presence of headings, etc.). These results clearly show document structure and referential strategies to vary according to text type. As Biber phrased it: "global generalizations are often not accurate at all, because there is no adequate overall linguistic characterization of the entire language; rather, there are marked linguistic differences across registers (or sublanguages). Thus a complete description of the language often entails a composite analysis of features as they function in various registers". (1993:220).

## REFERENCES

- ARIEL, M. 1990. *Accessing noun phrase antecedents*. London: Routledge.
- BESTGEN, Y., VONK, W. 2000. Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language* 42: 74-87.
- BESTGEN, Y., COSTERMAN, J. 1997. Temporal markers of narrative structure. Studies in production. In J. COSTERMANS, M. FAYOL (eds.), *Processing interclausal relationships in the production and comprehension of text*. Mahwah, NJ : Erlbaum. 201-218.
- BESTGEN, Y., DEGAND, L., SPOOREN, W. 2006. Toward Automatic Determination of the Semantics of Connectives in Large Newspaper Corpora. *Discourse Processes* 41(2): 175-193.
- BIBER, D. 1993. Using register-diversified corpora for genral language studies. *Computational Linguistics* 19(2): 219-241.
- BILHAUT, F., HO-DAC, M., BORILLO, A., CHARNOIS, T., ENJALBERT, P., DRAOULEC, A. L., MATHET, Y., MIGUET, H., PÉRY-WOODLEY, M.-P., SARDA, L. 2003. Indexation discursive pour la navigation intra-documentaire : cadres temporels et spatiaux dans l'information géographique. *Proceedings of Traitement Automatique du Langage Naturel (TALN)* : Batz-sur-Mer, France, 315-320.
- BOURIGAULT D. 2007. *Un analyseur syntaxique opérationnel : SYNTAX*, mémoire d'HDR en sciences du langage, CLLE-ERSS, Toulouse, France.
- CHAFE, W.L. 1976. Givenness Contrastiveness definiteness subjects topics and point of view'. In C.N.LI (ed) *Subject and Topic*. New York: Academic Press, 25-55.
- CHAROLLES, M. 1997. L'encadrement du discours; univers champs domaines et espaces. *Cahier de Recherche Linguistique* 6. Nancy: LanDisCo université Nancy2.
- CHAROLLES, M., LE DRAOULEC, A., PERY-WOODLEY, M.-P., SARDA, L. 2005. Temporal and spatial dimensions of discourse organisation. *Journal of French Language Studies* 15(2): 115-130.
- CHAROLLES, M., VIGIER, D. 2005. Les adverbiaux en position préverbale : portée cadrative et organisation des discours. *Langue Française* 148: 9-30.
- CROMPTON, P. 2006. The effect of position on the discourse scope of adverbials. *Text and Talk* 3: 245-279.

- DE MULDER, W. 1997. Les démonstratifs : des indices de changement de contexte. In N. FLAUX, D. VAN DE VELDE & W. DE MULDER (eds) *Entre général et particulier : les Déterminants*. Besançon: Artois Press Université, 137-200.
- DIK, S.C. 1997. *Theory of Functional Grammar: Complex and Derived Constructions*. Berlin & New-York: Mouton de Gruyter.
- FOSSARD, M. 2006. Aspects psycholinguistiques du traitement des démonstratifs: résultats croisés en français et en anglais. *Langue Française* 152: 82-95.
- GOUTSOS D. 1996. A model of sequential relations in expository text. *Text* 16(4): 501-533.
- GUNDEL, J. K., HEDBERG N. & ZACHARSKI R. 1993 Cognitive status and the form of referring expressions in discourse. *Language* 69(2): 274-307.
- HO-DAC, L.-M. 2008. *Discourse organisation through Theme position*. In S. Neumann and E. Steiner (eds.) Online Proceedings of the 19th European Systemic Functional Linguistics Conference and Workshop, Saarbrücken, 23-25 July 2007. URL: <http://scidok.sulb.uni-saarland.de/volltexte/2008/1691/>
- HO-DAC, L.-M. 2007. *Exploration en corpus de la position initiale dans l'organisation du discours*. Thèse de doctorat non publiée, Université de Toulouse le Mirail, Toulouse.
- HO-DAC, L.-M., PÉRY-WOODLEY, M.-P. 2008. Temporal adverbials and discourse segmentation revisited. *Linearisation and Segmentation in Discourse. Multidisciplinary Approaches to Discourse 2008 (MAD 08)* W. Ramm & C. Fabricius-Hansen (eds.). Oslo: Dept. of Literature, Area Studies and Europ. Languages, Univ. of Oslo. 65-76.
- LAMBRECHT, K. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.
- LE DRAOULEC, A., PÉRY-WOODLEY, M.-P. 2003. Time travel in text: Temporal framing in narratives and non-narratives. In L. LAGERWERF, W. SPOOREN, L. DEGAND (eds) *Determination of Information and Tenor in Texts : MAD 2003*. Amsterdam & Münster: Stichting Neerlandistiek & Nodus Publikationen, 267-275.
- LE DRAOULEC, A., PÉRY-WOODLEY, M.-P. 2005. Encadrement temporel et relations de discours. *Langue Française* 148: 45-60.
- LEVELT, W.J.M. 1981. The speaker's linearization problem'. *Philosophical Transactions of the Royal Society of London* B295: 305-315.
- PIERARD, S., BESTGEN, Y. 2006. Validation d'une méthodologie pour l'étude des marqueurs de la segmentation dans un grand corpus de textes. *TAL* 47(2): 89-110.
- POWER, R., SCOTT, D. & BOUAYAD-AGHA, N. 2003. Document Structure. *Computational Linguistics* 29(2): 211-260.
- SCHNEDECKER, C. 2005. Les chaînes de référence dans les portraits journalistiques : éléments de description. *Travaux de Linguistique* 5: 85-133.
- VIRTANEN, T. 1992. *Discourse Functions of Adverbial Placement in English : Clause-Initial Adverbials of Time and Place in Narratives and Procedural Place Descriptions*. Abo: Abo Akademi University Press.

VIRTANEN, T. 2004. Point of departure: Cognitive aspects of sentence-initial adverbials. In T. VIRTANEN (ed.) *Approaches to Cognition through Text and Discourse*. (Trends in Linguistics. Studies and Monographs 147.) Berlin/New York: Mouton de Gruyter, 79-97.

WALKER, M.A., JOSHI A. & PRINCE E. 1998. Centering in Naturally Occuring Discourse: An Overview. In M. Walker, A. Joshi & E. Prince (eds) *Centering Theory of Discourse*. Oxford: Calendron Press, 1-28.

## APENDIX

English glosses are provided here for examples [2], [4], [5], [6] and [7]. The position of temporal adverbials has been left as in the French texts, at the risk of making the translations ‘un-English’, in order to help comprehension of the points they illustrate. Temporal adverbials are in bold and relevant grammatical subject in bold italicized.

[2']

### 5. SECONDARY EDUCATION [heading]

**In the 1970's**, the problems of secondary schools made educational headlines: the numbers were too high, the courses were ill-adapted to the diversity of pupils, many were left on the wayside in an excessively elitist system. These difficulties have not disappeared, but the sixth-form (lycée) is where they are now most strongly felt.

#### 5.1 UNEQUAL GROWTH OF NUMBERS IN SECONDARY EDUCATION [subheading]

**At the start of the school year 1965-66**, there were 3,250,000 pupils in secondary education and sixth form colleges. Twenty years later, they were 5,100,000, and over 5,500,000 last year. [...]

**From 1965 to 1985**, the numbers of secondary school and sixth form students increased by 70%, but with different rhythms and levels of intensity depending on the school authority and the region. [...]

**At the start of the school year 1990-91**, the 6,700 secondary schools in mainland France took in 3,100,00 pupils, a decrease of over 300,000 compared with the school year 1985-86. [...]

[4']

### 1.3) THE PROFESSOR AND FOLKLORE SPECIALIST [heading]

**In January 1907**, *Béla Bartók* is appointed as a professor [...]. This appointment means he is free from material worries. **He** carries out his teaching duties very conscientiously. **He** follows up [...]. Having got over his passion for Strauss, **he** discovers Debussy's music [...]. The message [...]. 1907 is marked by a Concerto [...]. 1908 is the beginning [...]. Having courageously overcome through hard work a failed love affair, **Bartók** marries, in November 1910, his pupil Márta Ziegler. The couple honeymoon in Paris, [...]. **In March 1910**, **he** is in Paris again [...]. A week later, the young Quartet [...]. **At the end of May**, **Bartók** plays his Rhapsody [...] in Zurich [...]. **From 1912**, disheartened by Hungarian society's lack of interest in the new music, **Bartók** retires from musical life, [...]; **he** nevertheless continues [...]

[5']

### A POSSIBLE SCENARIO FOR A PREDICTABLE WAR [heading]

The plasticity of [Saddam Hussein's] regime is a factor which has been ignored in all the anticipations of the war [...]. Foreign attacks and various forms of intervention have prepared him for this ultimate confrontation. They have taught him to hide his targets [...]. The regime sometimes hides the most ordinary targets. **During the 1998 bombings, schools**, industrial sites and food warehouses became hiding places for ammunition stocks. **These schools** now shelter Party members in charge of maintaining order. These officials have left their official accommodation, [...].

[6']

**BOSNIA** [heading]

**In 1993**, the debates over the possibility of military engagement in Bosnia went together with numerous critical comments about the cost of operations, and about the need to involve the European allies' armed forces. Members of Congress asked questions about the war powers given to President Bill Clinton [...]. Some Congress members showed scepticism toward the pretensions of the White House to act as the "world's policeman". They considered that excessive intervention abroad would mean that their country was taking the risk of wasting its resources and of revealing its forces to its enemies. Senator John McCain (Republican, Arizona) thought for instance that, "by using our forces and our prestige in a thoughtless manner, we'll be wasting resources that we haven't got".

Similar comments were made **in 1995**, by Democratic members among other, judging that the Administration had paid too much attention to the aspirations of the conservative wing of Congress [...]

[7']

**1.1) YOUTH** [heading]

His childhood and adolescence took place in Italy, and later on the French Riviera and in Monaco, where he lived from 1886 to 1898 with his mother and his brother, born two years after him.

**In 1899, he** is in Paris with his family: a difficult existence, which does not prevent him from writing and studying in libraries. The summer of 1899 will see him spend a few weeks with his brother in Stavelot [...].

**In August 1901, he** sets off for the Rhineland where he is employed as a tutor. This year in Germany is decisive in the elaboration of his mental universe and in his love life. [...] he falls in love with his pupil's English governess, Annie Playden, who eventually rejects him after he had hoped that she would return his love: deeply shaken, he returns to France **in late August 1902**.

**In the course of that year, he** did more than live intensely. He wrote numerous poems, tales, chronicles, he absorbed images and themes which will stay with him. It is also during that year that the *Revue Blanche* of March 1902 publishes his first tale, "L'Hérésiarque", signed with the pen-name he had created for himself a long time previously on the basis of two of his first names: Guillaume Apollinaire.

*To quote this paper :*

e-reference: **Lydia-Mai Ho-Dac and Marie-Paule Péry-Woodley**, A data-driven study of temporal adverbials as discourse segmentation markers, *Discours*, 4 | 2009, [En ligne], mis en ligne le 30 juin 2009. URL : <http://discours.revues.org/index5952.html>.